Review of Ordinary Least Square & Maximum Likelihood

Why OLS is not good for insurance application

Week of 09/11/17

Part B

Linear model

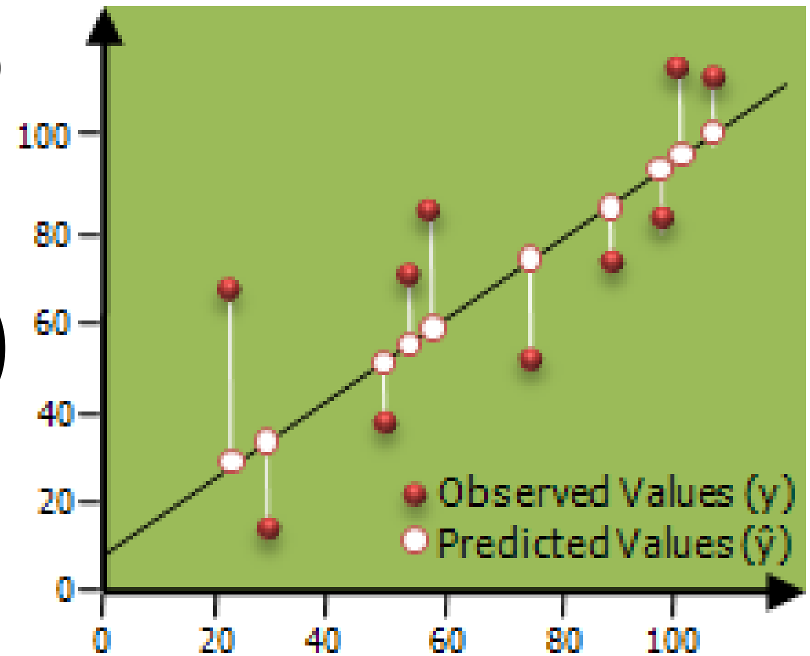
$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_n X_{i,n} + \varepsilon$$
$$= \Sigma_j \beta_j X_{i,j} + \varepsilon = \overline{\boldsymbol{X}}\overline{\boldsymbol{\beta}} + \varepsilon$$

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ simple linear model

Ordinary least squares

assumption: $\varepsilon \sim N(0, \sigma^2)$

$$\hat{\beta} = arg\,min(RSS)$$
$$= arg\,min(\Sigma_i(\widehat{y_i} - y_i)^2)$$
$$= arg\,min\left(\Sigma_i\left(\Sigma_j \beta_j X_{ij} - y_i\right)^2\right)$$
$$= (X'X)^{-1}X'Y$$



For simple linear model:

$$\widehat{\beta_1} = (x_i y_i - \frac{1}{n} x_i y_i)/(x_i^2 - \frac{1}{n}(x_i)^2) = \rho_{xy}$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

Hat Matrix

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$$

$$H = X(X'X)^{-1}X'$$

idempotent matrix $A^2 = AA = A$

Residual

$$\hat{e} = y - X\hat{\beta} = (I - H)y$$

Total $TSS = \sum_i (Y_i - \bar{Y})^2$

Regression $\mathrm{R}SS = \sum_i (\widehat{Y}_i - \bar{Y})^2$

Error $\mathrm{E}SS = \sum_i (Y_i - \widehat{Y}_i)^2$

We should have $TSS = ESS + RSS$

$$
\begin{aligned}
TSS = \sum (Y_i - \bar{Y})^2 &= \sum \left[\left(Y_i - \widehat{Y}_i \right) + (\widehat{Y}_i - \bar{Y}) \right]^2 \\
&= RSS + ESS + 2 \sum (Y_i - \widehat{Y}_i)(\widehat{Y}_i - \bar{Y}) \\
&= RSS + ESS + 2 \left[\sum \widehat{Y}_i \left(Y_i - \widehat{Y}_i \right) - \sum \bar{Y} \left(Y_i - \widehat{Y}_i \right) \right] \\
&= RSS + ESS + 2 \left[\sum \widehat{Y}_i e_i - \bar{Y} \sum e_i \right] \\
&= RSS + ESS
\end{aligned}
$$

($\widehat{Y}_i = \sum_j \beta_j X_{i,j}$ and $\sum_i e_i X_{i,j} = 0$)

So $$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

- RSS/TSS is the portion that has been explained by regression model
- ESS/TSS is the portion of TSS for the error

R square $$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \frac{\sum_i (Y_i - \widehat{Y_i})^2}{\sum_i (Y_i - \bar{Y})^2}$$

- Why this is biased?

Adjust R square

$$R_{adj}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} = \frac{(n-1)R^2 - (k-1)}{n-k}.$$

Likelihood function

$$L(\theta|x) = \prod p_\theta(x) \quad or \prod f_\theta(x)$$

MLE: $\hat{\theta} = \underset{\theta}{\text{arg max}}\, L(\theta|x) = \underset{\theta}{\text{arg min}} [-\ln(L(\theta|x))]$

Score function $U(\theta) = \frac{\partial \ln(L(\theta|x))}{\partial\theta}$

Observed information $I(\theta) = -\frac{\partial U(\theta)}{\partial\theta} = -\frac{d^2 \ln(L(\theta|x))}{d\theta^2}$

Find $\hat{\theta}$ => find minimum $-\ln(L(\theta|x))$

- Iterative numerical techniques

  $\theta_{n+1}=\theta_n - L(\dot{\theta})/L(\ddot{\theta})$, similar to Newton's method $x_{n+1}=x_n - f(x_n)/f'(x_n)$

- Use statistical analysis application, such as *R*

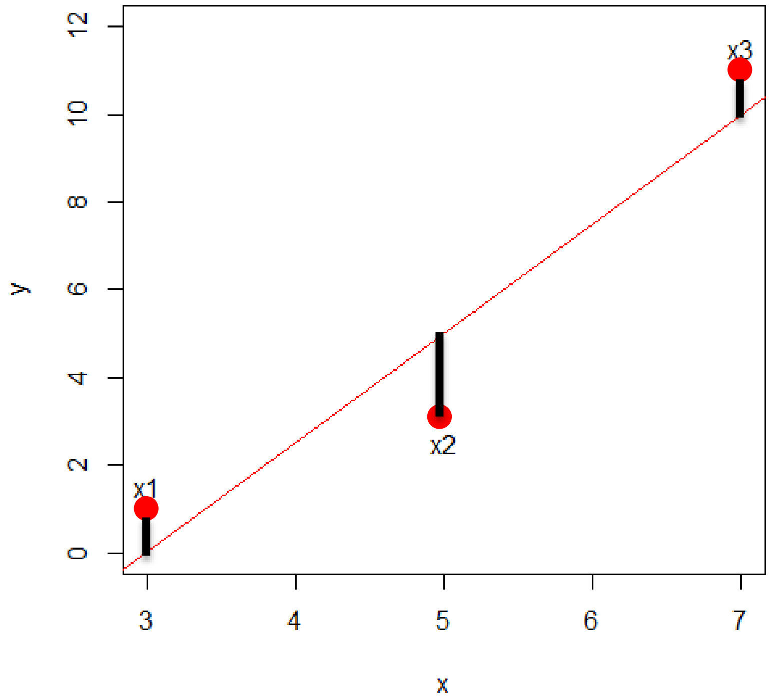
OLS and MLE are equivalent for normal

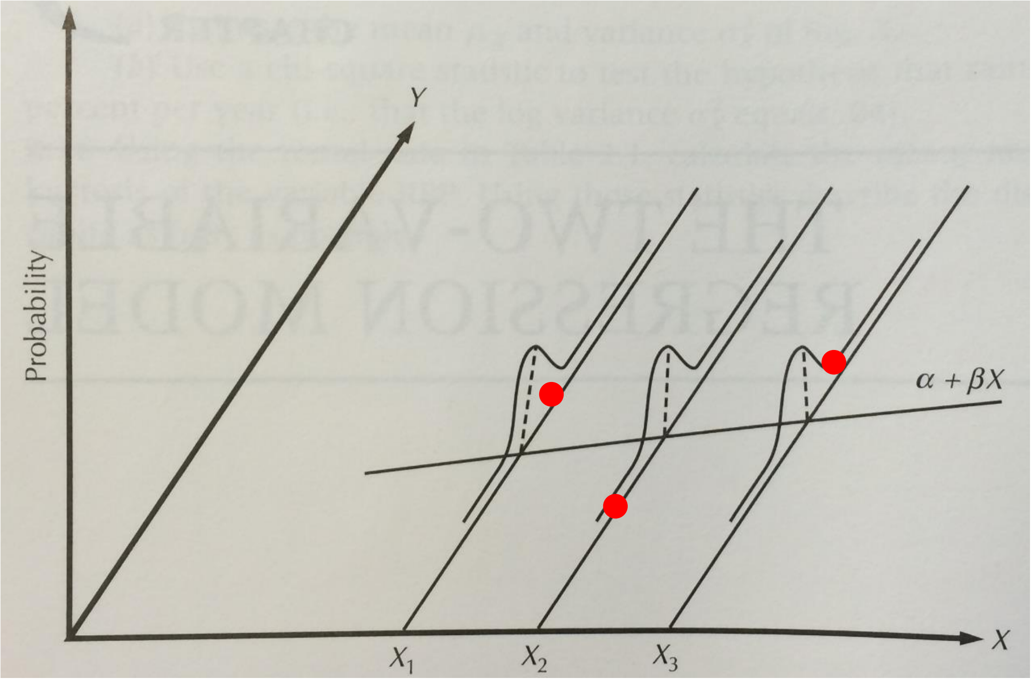
OLS minimize $\sum_{i=0}^{n}(Y_i - \widehat{Y}_i)^2$

MLE $\max\left(\mathrm{L}(\theta|\mathrm{X})\right) = \min\left(-\ln\left(\mathrm{L}(\theta|\mathrm{X})\right)\right)$

$$\mathrm{L}(\theta|\mathrm{X}) \sim \exp\left(-\sum_{i=0}^{n}(Y_i - \mu_i)^2 / 2\sigma^2\right)$$

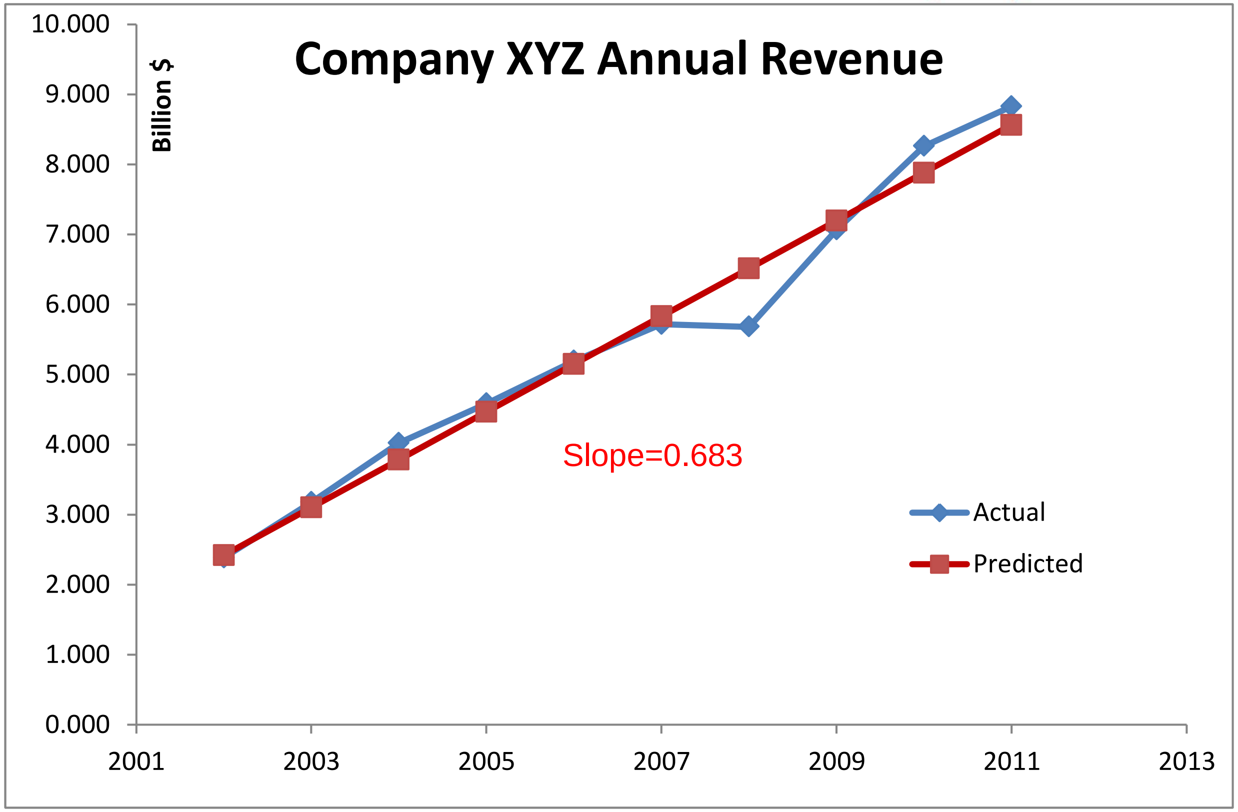
$$\mu_i = \widehat{Y}_i = E(Y_i) = \sum_{j=1}^{m} x_{i,j}\beta_j = X\beta$$





Example #1

Year	Revenue
2002	2.382
2003	3.175
2004	4.021
2005	4.585
2006	5.194
2007	5.718
2008	5.681
2009	7.067
2010	8.262
2011	8.830

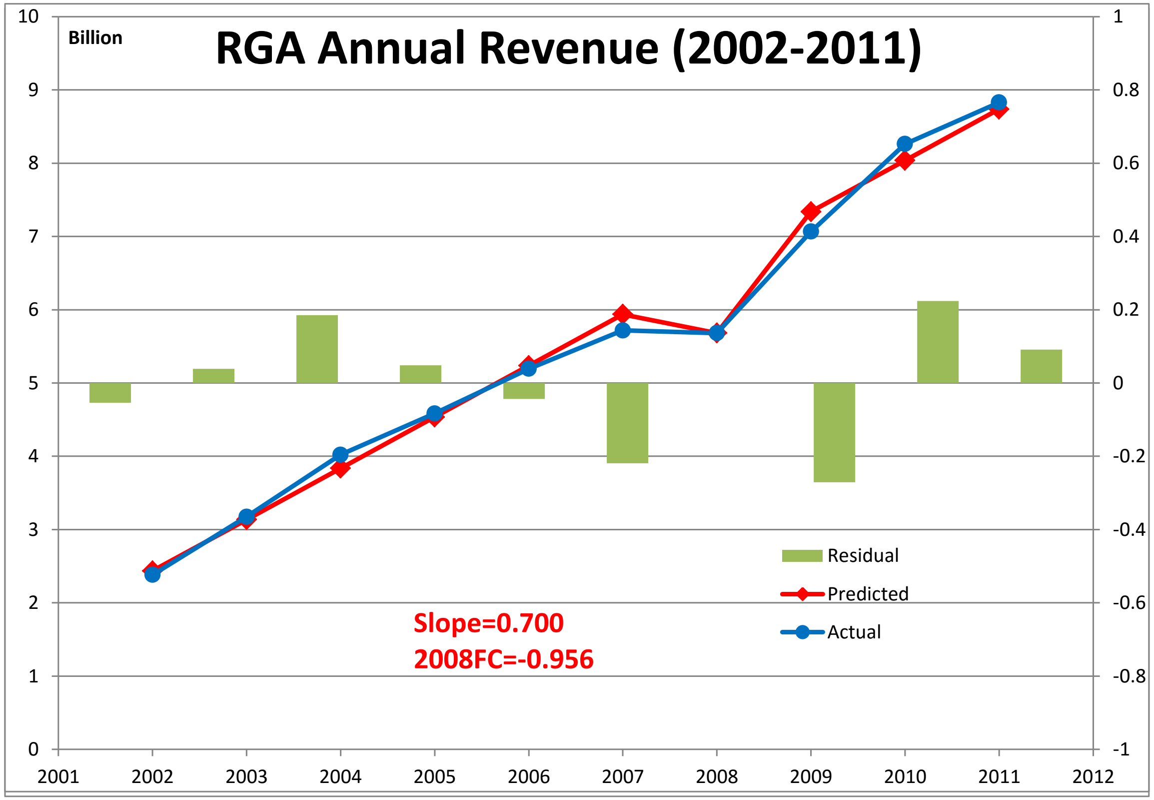


In R

```
> iData<-read.table("iData1.txt", header = TRUE, sep=",")
> iModel <- lm(Revenue~Year, data=iData)
> summary(iModel)
```

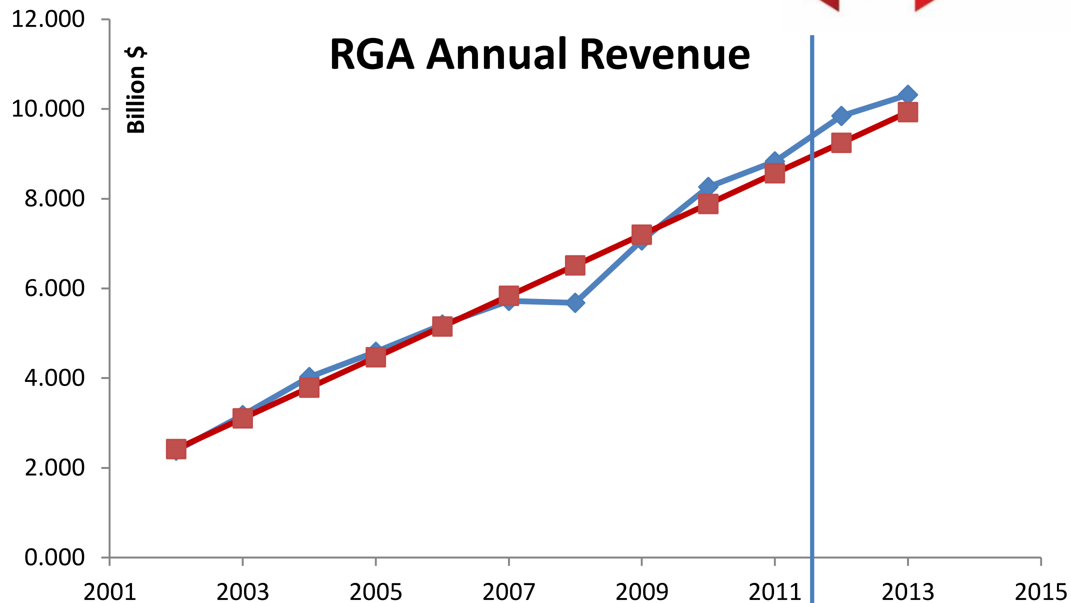
- Dummy variable
  - Binary, indicator, dichotomous, discrete, categorical variable
  - Absence or presence; base or effect
  - Convert category variable to numeric
  - Seasonality study, or
- 2008 financial crisis

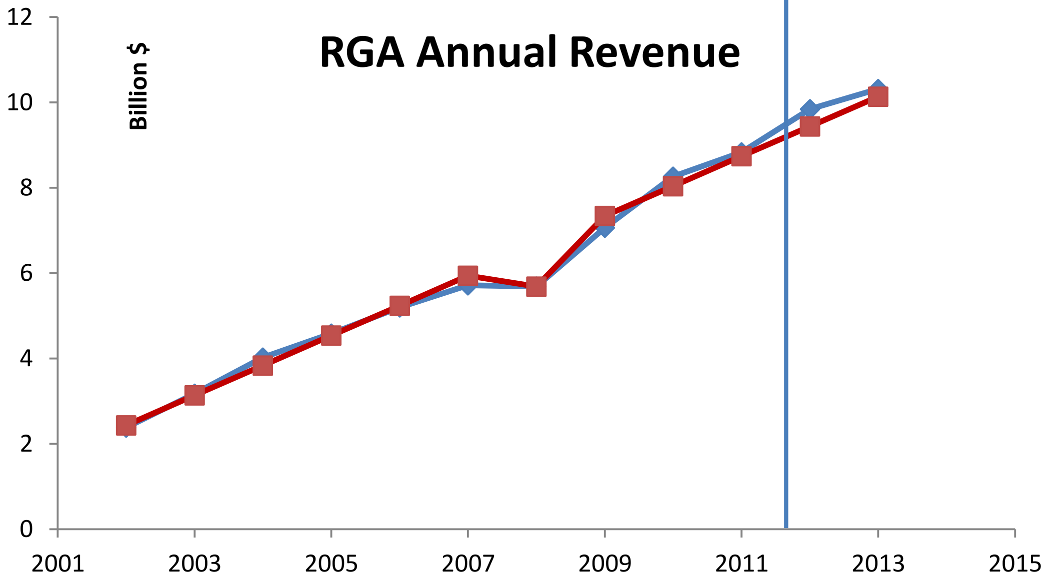
Year	2008FC	Revenue
2002	0	2.382
2003	0	3.175
2004	0	4.021
2005	0	4.585
2006	0	5.194
2007	0	5.718
2008	**1**	5.681
2009	0	7.067
2010	0	8.262
2011	0	8.830



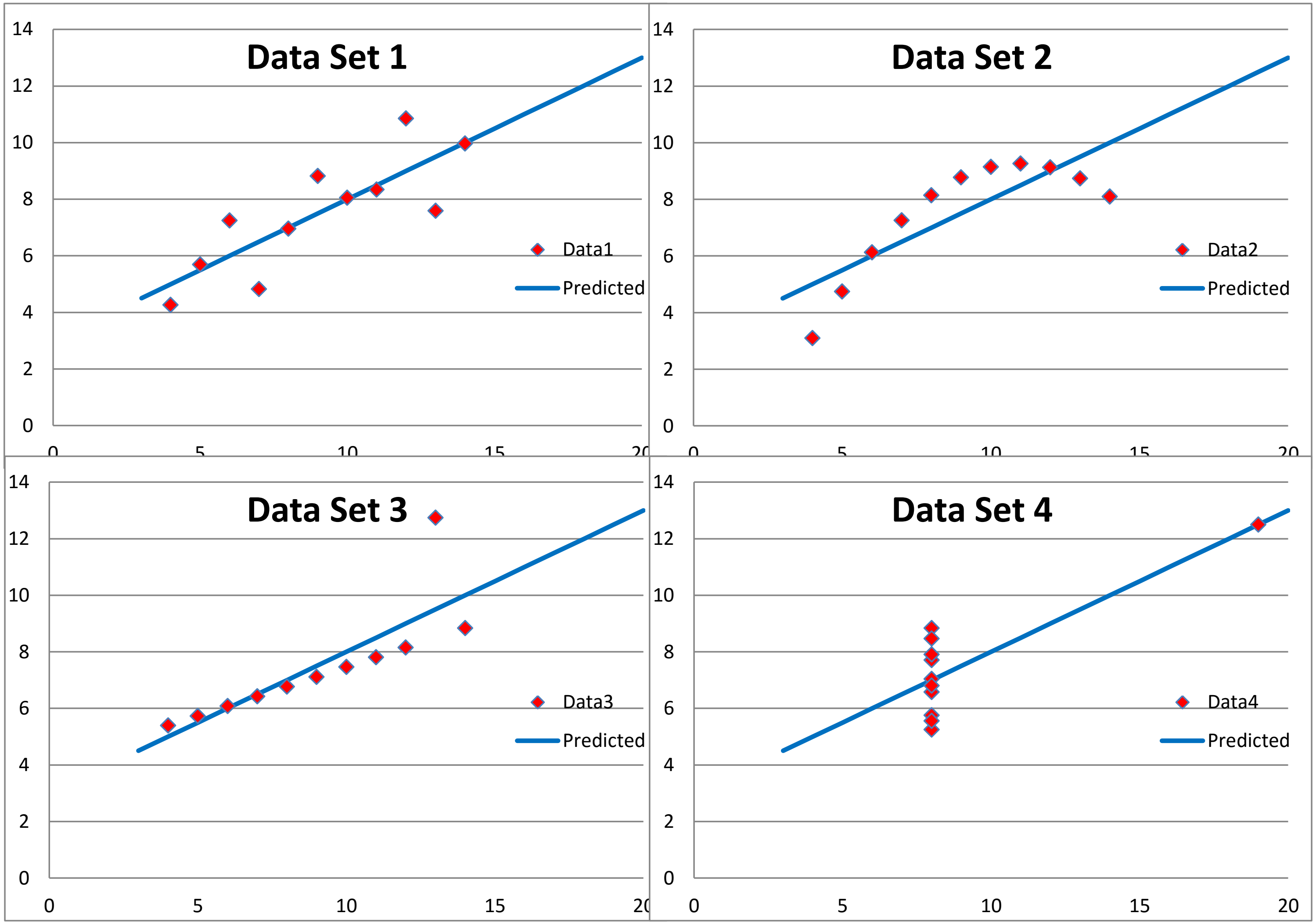
Prediction

Year	Revenue
2002	2.382
2003	3.175
2004	4.021
2005	4.585
2006	5.194
2007	5.718
2008	5.681
2009	7.067
2010	8.262
2011	8.830
2012	9.841
2013	10.318





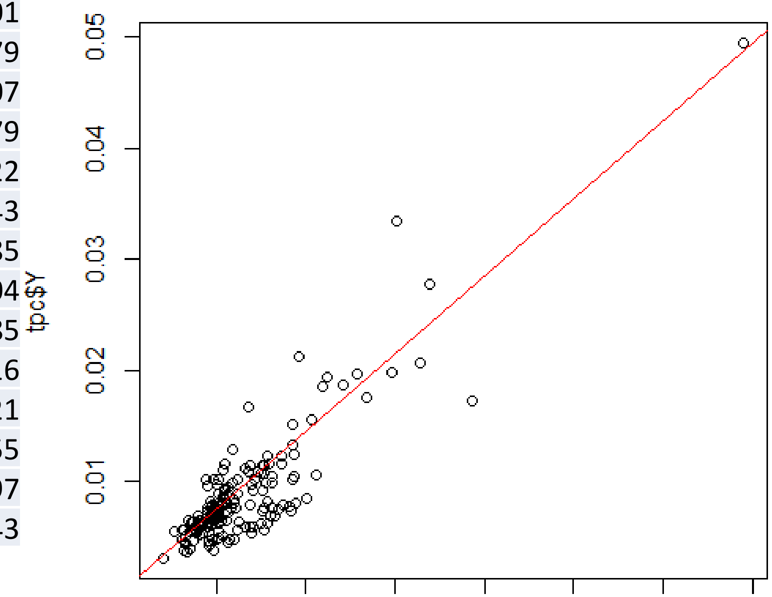
Understanding Data



Conclusion: you can not apply model without understanding data

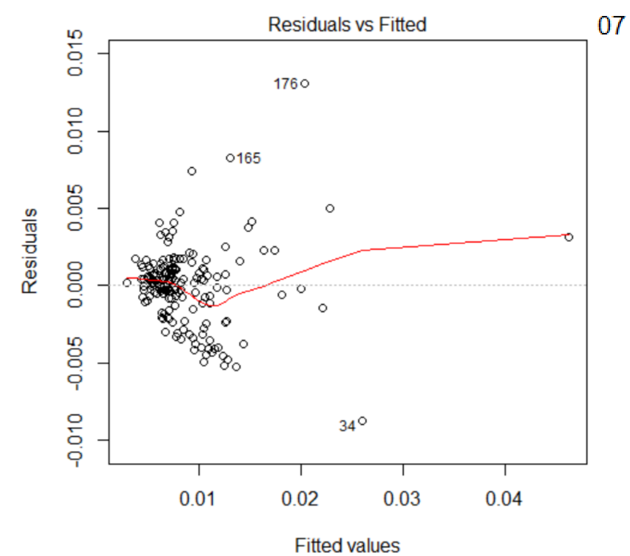
Example #2

lga	sd	claims	accidents	ki	population	pop_density
ASHFIELD	1	1103	2304	920	124850	0.499001
AUBURN	1	1939	2660	1465	143500	0.148379
BANKSTOWN	1	4339	7381	3864	470700	0.205407
BAULKHAMHILLS	1	1491	3217	1554	311300	0.025879
BLACKTOWN	1	3801	6655	4175	584900	0.081222
BOTANY	1	387	2013	854	106350	0.178143
BURWOOD	1	1299	1888	946	88750	0.413285
CAMDEN	1	326	510	353	56900	0.009404
CAMPBELLTOWN	1	1789	2872	1846	354050	0.03785
CANTERBURY	1	3310	4991	2497	396100	0.392816
CONCORD	1	443	1182	497	71200	0.204721
DRUMMOYNE	1	673	1340	517	95950	0.39755
FAIRFIELD	1	6524	6977	3332	457150	0.150497
HOLROYD	1	1217	3672	1526	243900	0.202343



Y= ki / population

X= accidents / population



Linear model (OLS) is very popular in almost every field

But actuarial science is exceptional, why?

Assumptions of OSL

a) The observations are independent.
b) Each observation has a normal distribution
c) The mean of an observation is a linear function of a set of explanatory variables.
d) The variance of observations is the same, regardless of the values of the explanatory variables

Questions:

a) If an actuary is modeling claim counts, which of the assumptions is clearly violated?

b) If Poisson distrib. is selected for claim counts, will the constant variance assumption hold? How about negative binomial?

c) Suppose you are modeling losses based on factors such as age, gender, & location. Why might a linear relationship not be appropriate? Other alternative?

Distribution	Mean	Variance	Sample Application
Normal	μ	1 (σ^2)	General Application
Poisson	μ	μ	Claim Frequency, Counts
Bernoulli	μ	$\mu(1-\mu)$	Retention, cross-sell, UW, rates
Negative binomial	μ	$\mu(1+\kappa\mu)$	Claim severity
Gamma	μ	μ^2/ν	Claim severity
Tweedie	μ	μ^p, $p\in(1,2)$	Claim Cost
Inverse Gaussian	μ	$\sigma^2\mu^3$	Claim severity